

# Acceleration on sparse promoting recovery and its benefits in seismic applications

Lina Miao, Felix Herrmann

---

**SLIM** 

University of British Columbia

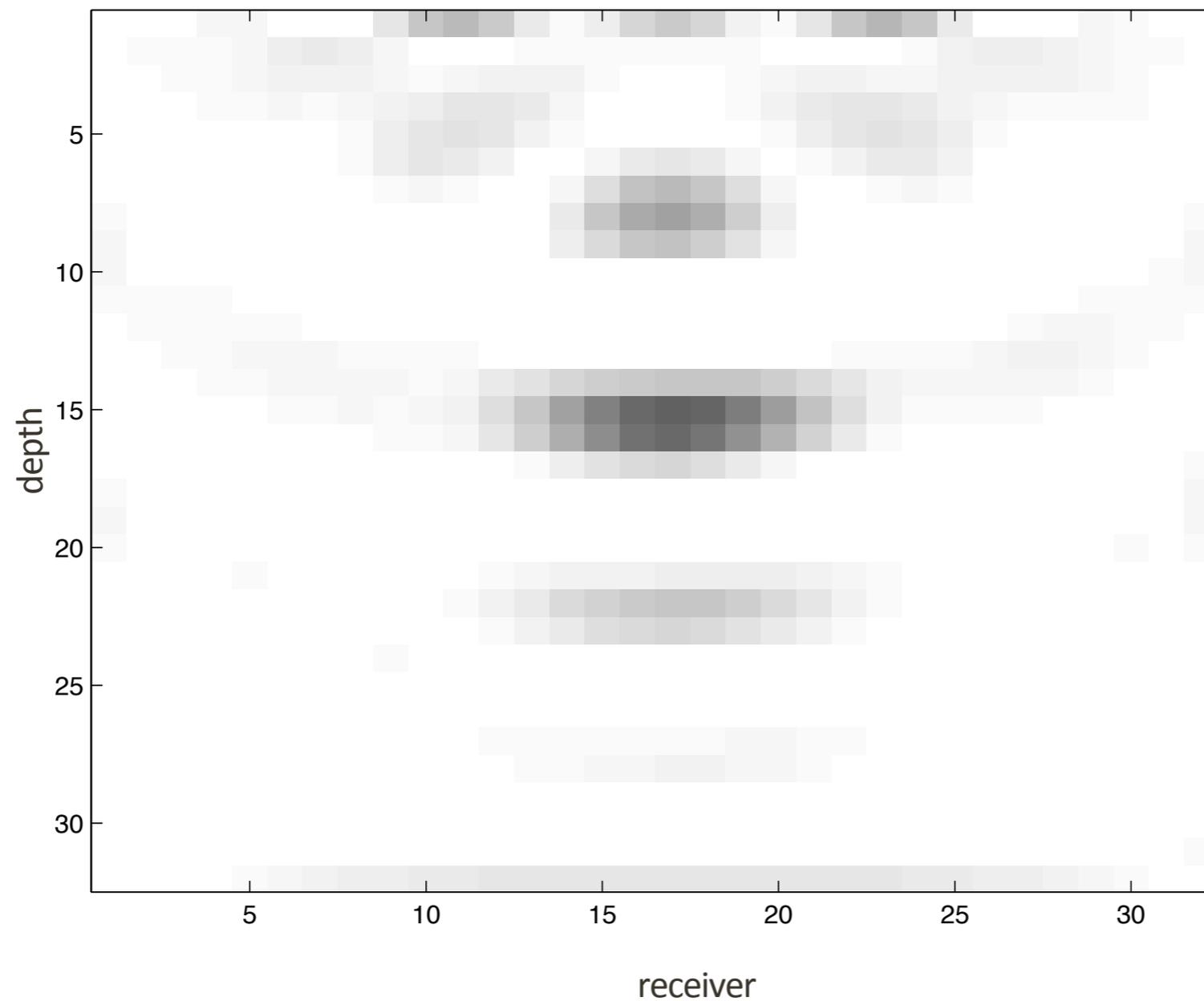
# Motivation

Iter	Objective	Relative Gap	Rel Error	gNorm	stepG	nnzX	nnzg
46	6.0646366e-01	3.1622007e+00	5.21e-01	4.347e+01	-0.3	0	0
47	6.0392961e-01	2.4731485e+00	5.19e-01	3.462e+01	0.0	0	0
48	5.9950646e-01	7.9860271e-02	5.14e-01	1.371e+00	0.0	0	0
49	5.9927042e-01	7.8260438e-02	5.14e-01	1.354e+00	0.0	0	0
50	5.9232993e-01	1.6187001e-01	5.07e-01	2.537e+00	0.0	0	0

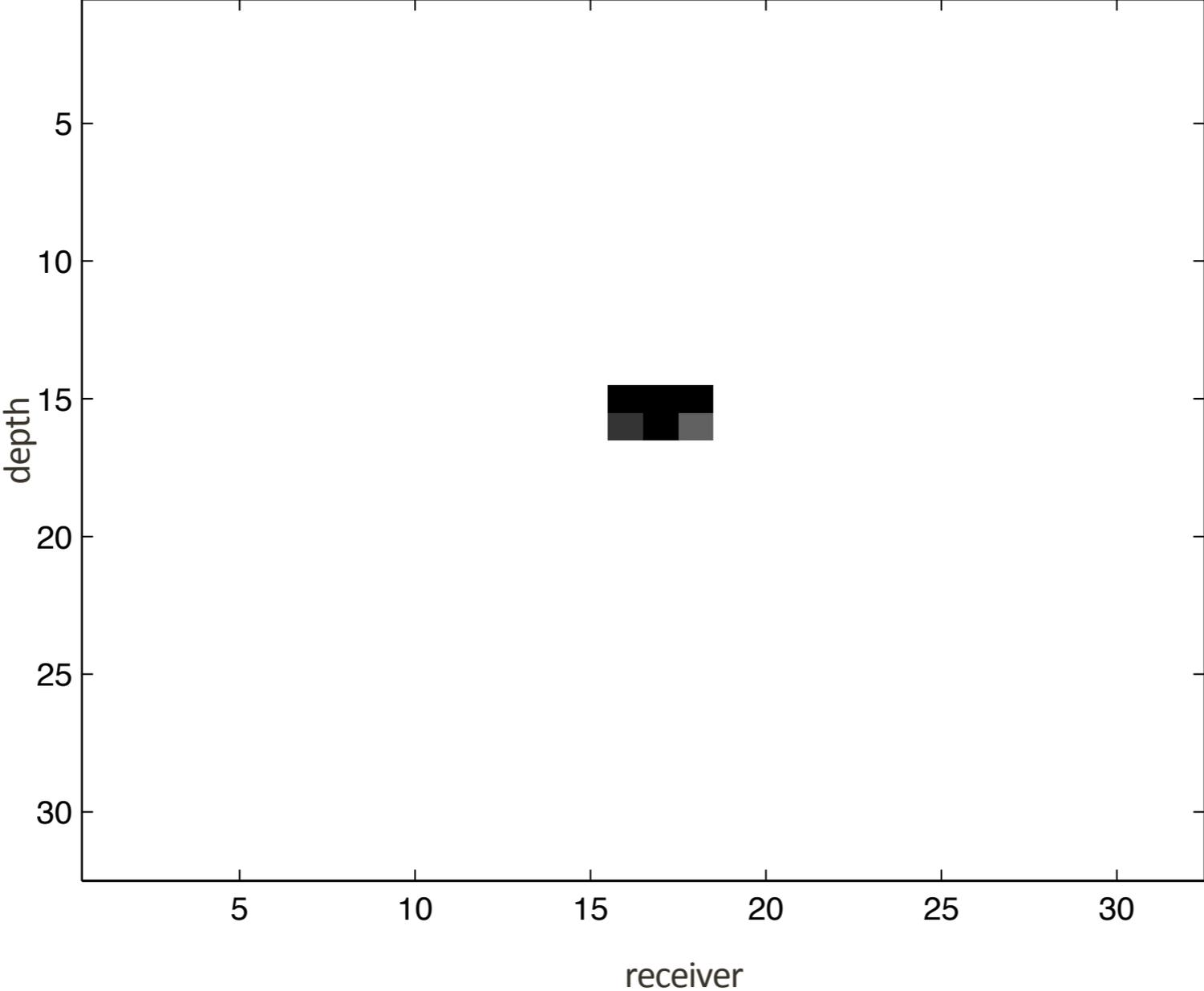
ERROR EXIT -- Too many iterations

Products with A	:	72	Total time (secs)	:	399.9
Products with A'	:	52	Project time (secs)	:	0.5
Newton iterations	:	2	Mat-vec time (secs)	:	398.4
Line search its	:	30	Subspace iterations	:	0

# Seismic Reverse Time Migration (Point scattering example)



# SPGI1 Recovery



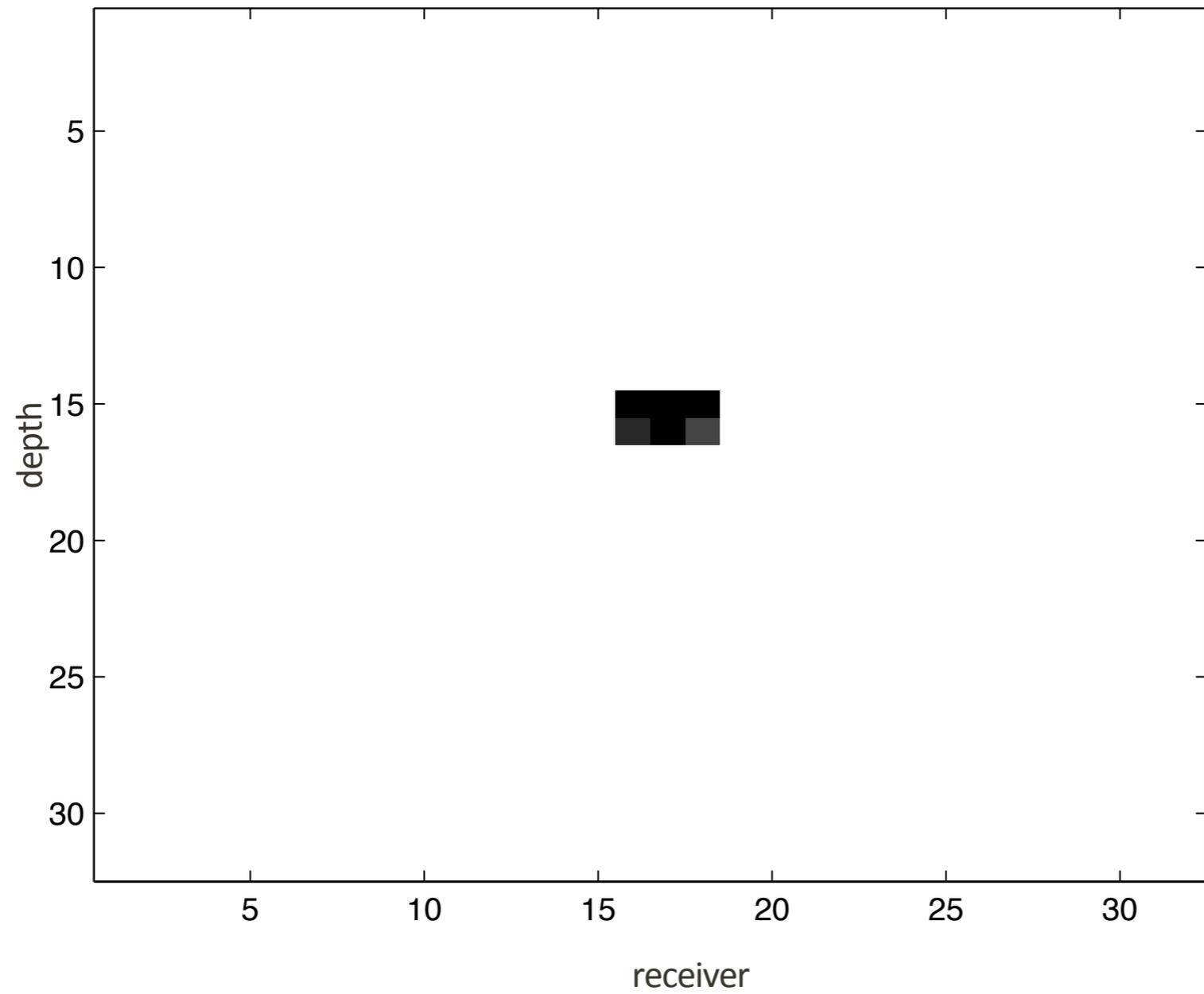
# SPG11 Recovery

Iter	Objective	Relative Gap	Rel Error	gNorm	stepG	nnzX	nnzg
68	2.2128499e+01	1.8467691e+00	1.00e+00	4.035e+04	0.0	10	0
69	2.2434980e+01	3.2091437e+01	1.00e+00	2.104e+05	-0.3	10	0
70	2.0834995e+01	1.7956228e+00	1.00e+00	4.252e+04	-0.3	10	0
71	1.9765301e+01	1.8445187e+00	1.00e+00	4.002e+04	0.0	10	0
72	1.9288837e+01	1.8069763e+00	1.00e+00	3.813e+04	0.0	10	0
73	1.9037702e+01	1.8294152e+00	1.00e+00	3.755e+04	0.0	10	0
74	1.8639298e+01	1.8310105e+00	1.00e+00	3.610e+04	0.0	10	0
75	1.7572991e+01	1.8291278e+00	1.00e+00	3.175e+04	0.0	10	0
76	1.7711170e+01	1.2603861e+01	1.00e+00	7.205e+04	0.0	10	0
77	1.7450281e+01	1.9027312e+00	1.00e+00	3.451e+04	-0.3	10	0
78	1.6903953e+01	1.8902180e+00	1.00e+00	3.290e+04	0.0	10	0
79	1.6545707e+01	1.8380806e+00	1.00e+00	3.113e+04	0.0	10	0
80	1.6396725e+01	1.8623832e+00	1.00e+00	3.128e+04	0.0	10	0

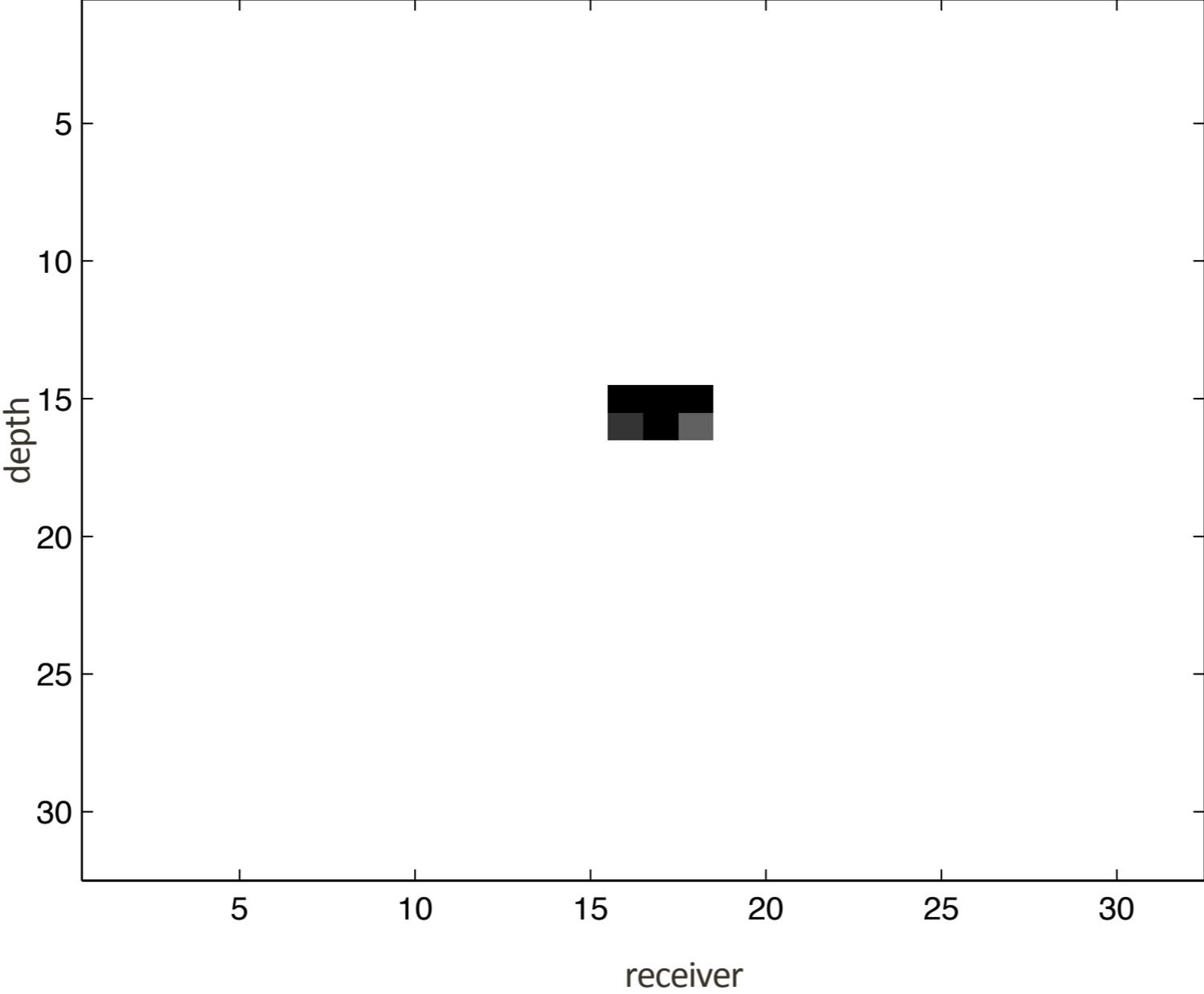
ERROR EXIT -- Too many iterations

Products with A	:	98	Total time (secs)	:	1094.7
Products with A'	:	81	Project time (secs)	:	0.5
Newton iterations	:	1	Mat-vec time (secs)	:	986.7
Line search its	:	22	Subspace iterations	:	0

# PQNI1 Recovery



# SPGI1 Recovery



# PQNI1 Recovery

Iteration	FunEvals	Projections	Step Length	rNorm2	opt Cond
5	1	71	1.00000e+00	4.86441e+01	8.71044e-02
6	1	92	1.00000e+00	4.60130e+01	8.67556e-02
7	1	130	1.00000e+00	4.11350e+01	8.58883e-02
8	1	158	1.00000e+00	3.40699e+01	8.42295e-02
9	1	180	1.00000e+00	2.99819e+01	8.28857e-02
10	1	197	1.00000e+00	2.69776e+01	8.16233e-02
11	1	214	1.00000e+00	2.34853e+01	8.01542e-02
12	1	224	1.00000e+00	2.24448e+01	7.96595e-02
13	1	233	1.00000e+00	2.19325e+01	7.93629e-02
14	1	243	1.00000e+00	1.88247e+01	7.97001e-02
15	1	254	1.00000e+00	1.84318e+01	7.94062e-02
16	1	265	1.00000e+00	1.70312e+01	7.81970e-02
17	1	278	1.00000e+00	1.50816e+01	7.63327e-02

rNorm less than sigma\_ref

EXIT - residual less than sigma ref

Products with A	:	20	Total time (secs)	:	255.9
Products with A'	:	20	Project time (secs)	:	4.5
Newton iterations	:	1	Mat-vec time (secs)	:	216.6

# SPG11 Recovery

Iter	Objective	Relative Gap	Rel Error gNotm	stepG	nnzX	nnzg
68	2.2128499e+01	1.8467691e+00	1.00e+00	4.035e+04	0.0	10
69	2.2434980e+01	3.2091437e+01	1.00e+00	2.104e+05	-0.3	10
70	2.0834995e+01	1.7956228e+00	1.00e+00	4.252e+04	-0.3	10
71	1.9765301e+01	1.8445187e+00	1.00e+00	4.002e+04	0.0	10
72	1.9288837e+01	1.8069763e+00	1.00e+00	3.813e+04	0.0	10
73	1.9037702e+01	1.8294152e+00	1.00e+00	3.755e+04	0.0	10
74	1.8639298e+01	1.8310105e+00	1.00e+00	3.610e+04	0.0	10
75	1.7572991e+01	1.8291278e+00	1.00e+00	3.175e+04	0.0	10
76	1.7711170e+01	1.2603861e+01	1.00e+00	7.205e+04	0.0	10
77	1.7450281e+01	1.9027312e+00	1.00e+00	3.451e+04	-0.3	10
78	1.6903953e+01	1.8902180e+00	1.00e+00	3.290e+04	0.0	10
79	1.6545707e+01	1.8380806e+00	1.00e+00	3.113e+04	0.0	10
80	1.6396725e+01	1.8623832e+00	1.00e+00	3.128e+04	0.0	10

ERROR EXIT -- Too many iterations

Products with A	:	98	Total time (secs)	:	1094.7
Products with A'	:	81	Project time (secs)	:	0.5
Newton iterations	:	1	Mat-vec time (secs)	:	986.7
Line search its	:	22	Subspace iterations	:	0

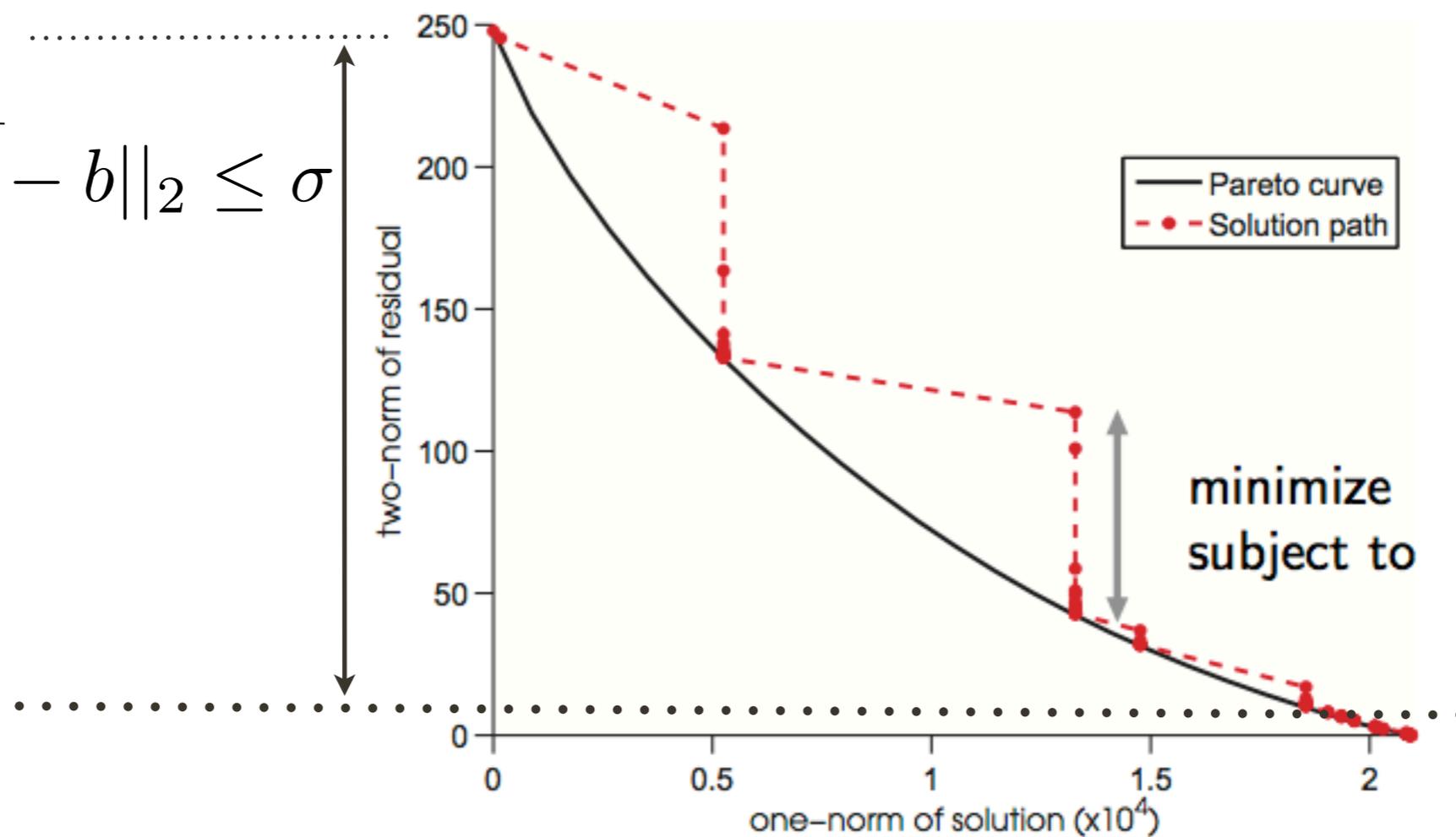
# Outline

---

- SPGI1 overview
- Acceleration with PQNI1
- Experiments
- Conclusion
- Future Plans

# SPG1 Solution Path

minimize  $\|x\|_1$   
 subject to  $\|Ax - b\|_2 \leq \sigma$



minimize  $\|Ax - b\|_2$   
 subject to  $\|x\|_1 \leq \tau$

solve with SPG

(spectral projected gradients)

# SPG vs PQN

- Spectral Projected Gradient
  - Gradient -- first order optimization method
  - Projected -- constrained optimization
  - Spectral -- line search technique
- Projected Quasi Newton
  - Newton -- second order optimization method
  - Quasi -- not quite exactly Newton
  - Projected -- constrained optimization

# SPG vs PQN

$$q_k(x) = f_k + (x - x_k)^T \nabla f(x_k) + 1/2(x - x_k)^T B_k(x - x_k)$$

First order information

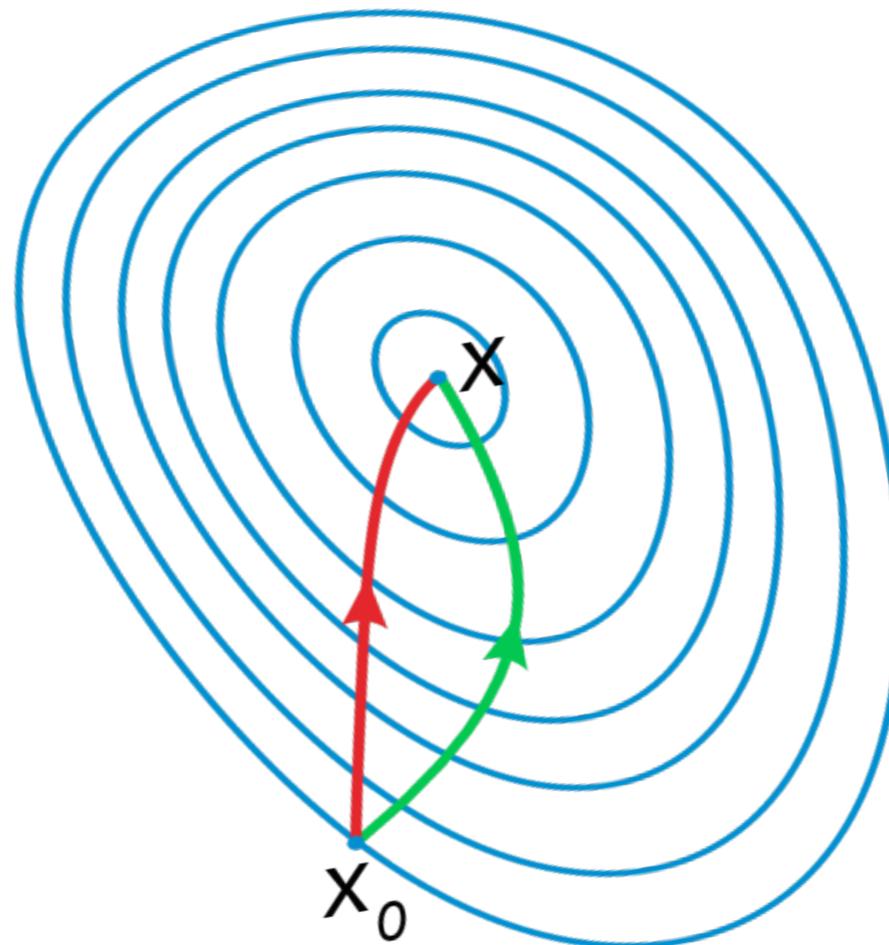


Second order information



# SPG vs PQN

$$q_k(x) = f_k + (x - x_k)^T \nabla f(x_k) + 1/2(x - x_k)^T B_k(x - x_k)$$



# SPG vs PQN

Objective :  $\min_{x \in C} f(x)$

**for**  $k = 1$  to  $iter_{SPG}$  **do**

$$x^{k+1} = \underset{x \in C}{\operatorname{argmin}} f(x^k) + (x - x^k)^T \nabla f(x^k) + \frac{1}{2\alpha} \|(x - x^k)\|_2^2$$

SPG

Closed form solution :

$$x^{k+1} = P_{x \in C}(x^k - \alpha \nabla f(x^k));$$

**end for**

**for**  $K = 1$  to  $iter_{PQN}$  **do**

$$x^{K+1} = \underset{x \in C}{\operatorname{argmin}} f(x^K) + (x - x^K)^T \nabla f(x^K) + \frac{1}{2} (x - x^K)^T B^K (x - x^K)$$

PQN

Does not have a closed form solution, solved via SPG approach:

**for**  $k = 1$  to  $iter_{SPG}$  **do**

...

**end for**

**end for**

# Experiments

- General stylized numerical example
- Usage and benefits in seismic data acquisition, processing, and inversion

# Numerical Example

- Basic compressed sensing example
  - Well conditioned
  - Random gaussian matrix
- Ill conditioned example

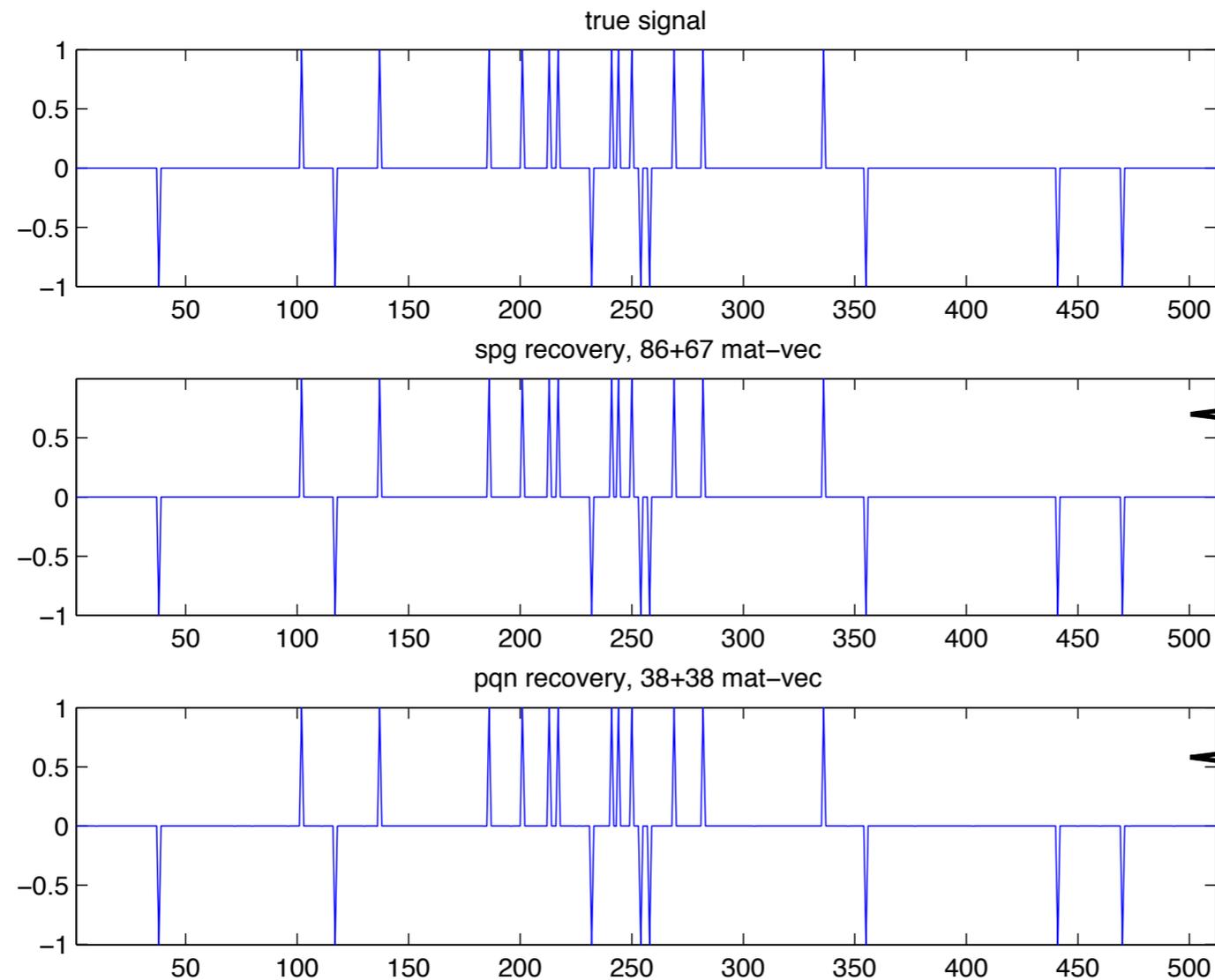
# SPGI1 vs PQNI1

(compressed sensing example)

- Recover a vector of length **512** components, of which **20** are nonzero, from **120** observations
- Sensing matrix is i.i.d Gaussian

# SPGL1 vs PQNL1

(compressed sensing example)

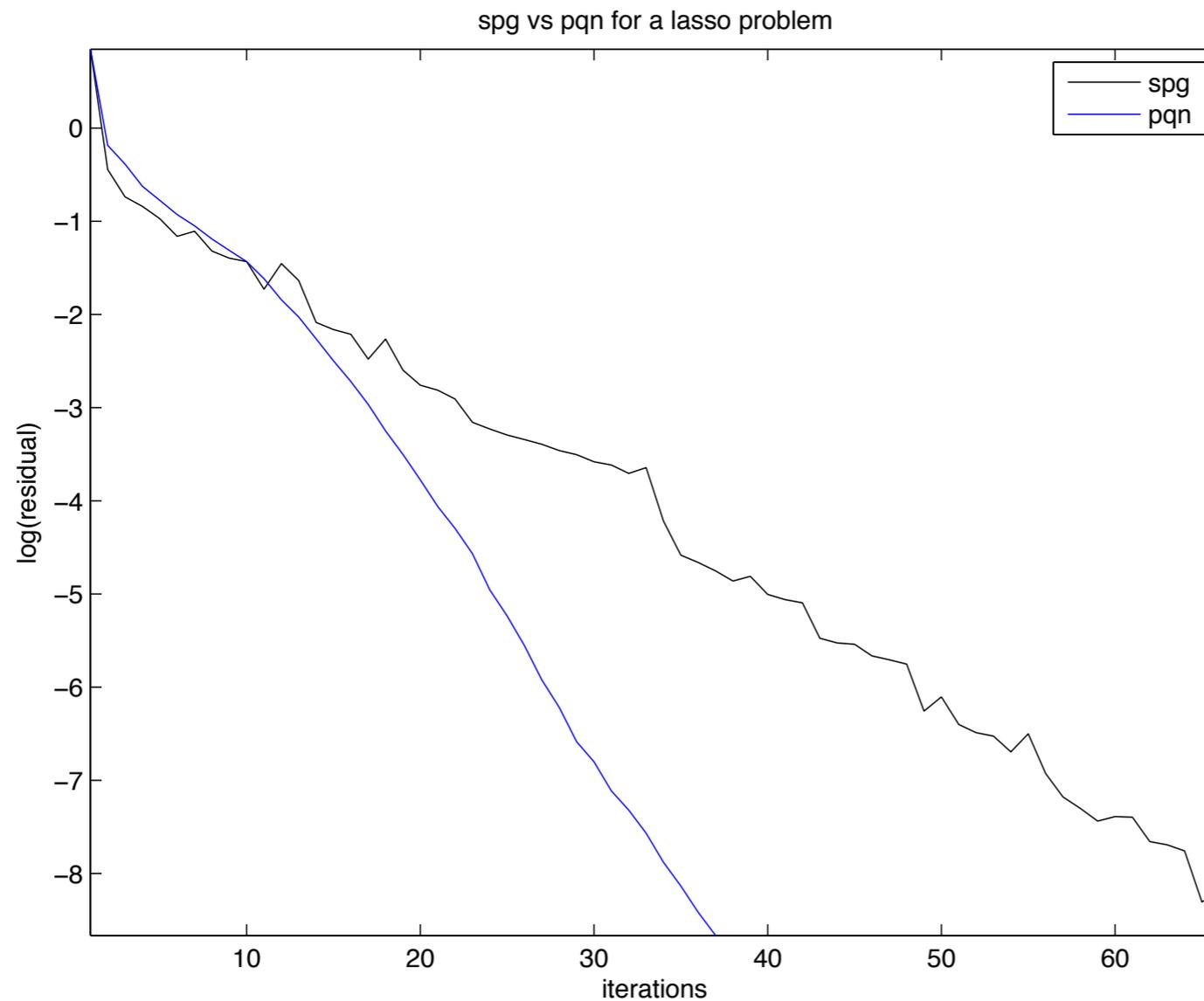


SPGL1 :  
86 + 67

PQNL1 :  
38 + 38

# SPGI1 vs PQNI1

(compressed sensing example)



# SPG vs PQN

Objective :  $\min_{x \in C} f(x)$

**for**  $k = 1$  to  $iter_{SPG}$  **do**

**SPG**

$$x^{k+1} = \underset{x \in C}{\operatorname{argmin}} f(x^k) + (x - x^k)^T \nabla f(x^k) + \frac{1}{2\alpha} \|(x - x^k)\|_2^2$$

Closed form solution :

$$x^{k+1} = P_{x \in C}(x^k - \alpha \nabla f(x^k));$$

**end for**

**for**  $K = 1$  to  $iter_{PQN}$  **do**

$$x^{K+1} = \underset{x \in C}{\operatorname{argmin}} f(x^K) + (x - x^K)^T \nabla f(x^K) + \frac{1}{2} (x - x^K)^T B^K (x - x^K)$$

**PQN**

Does not have a closed form solution, solved via SPG approach:

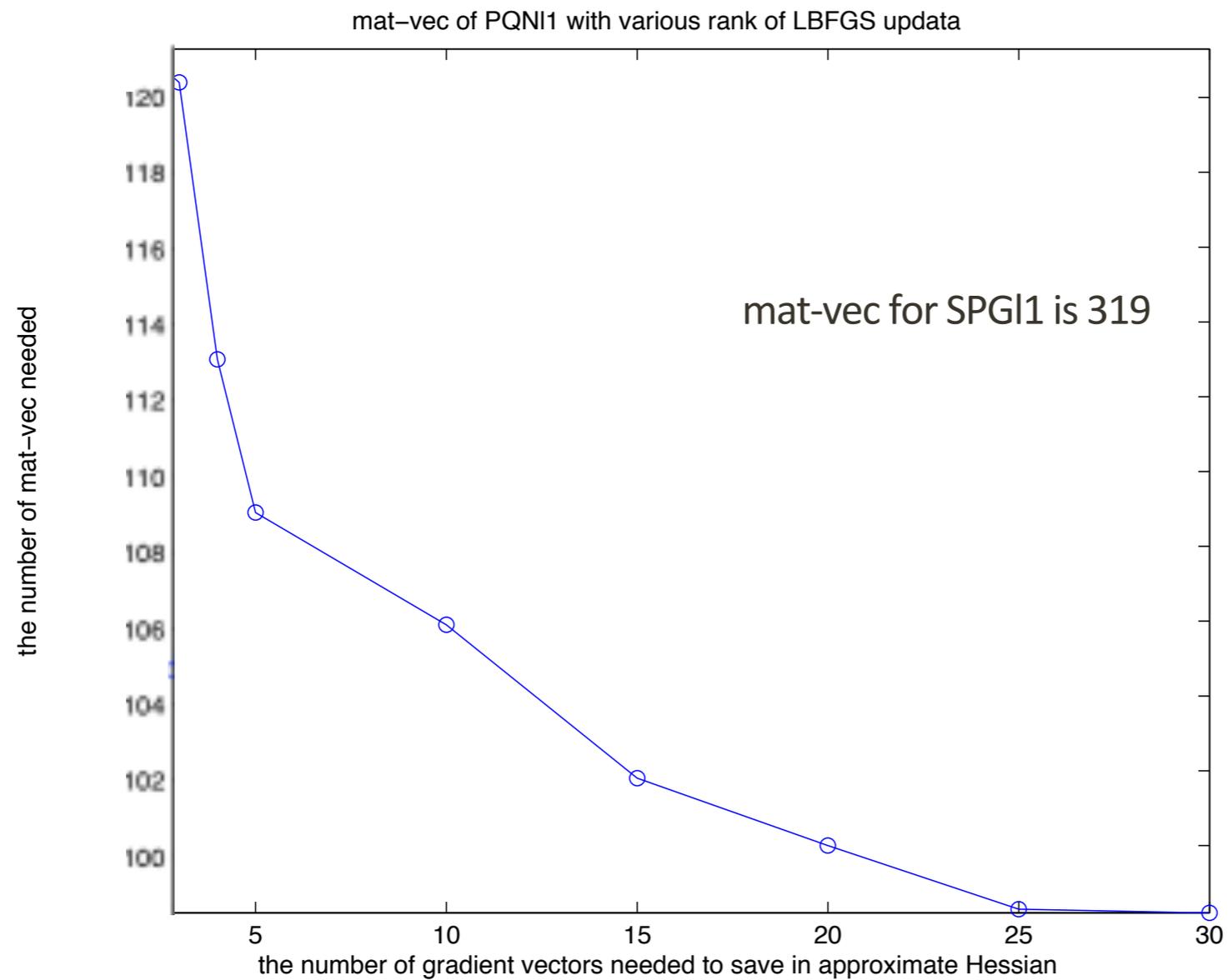
**for**  $k = 1$  to  $iter_{SPG}$  **do**

...

**end for**

**end for**

# Memory vs CPU time



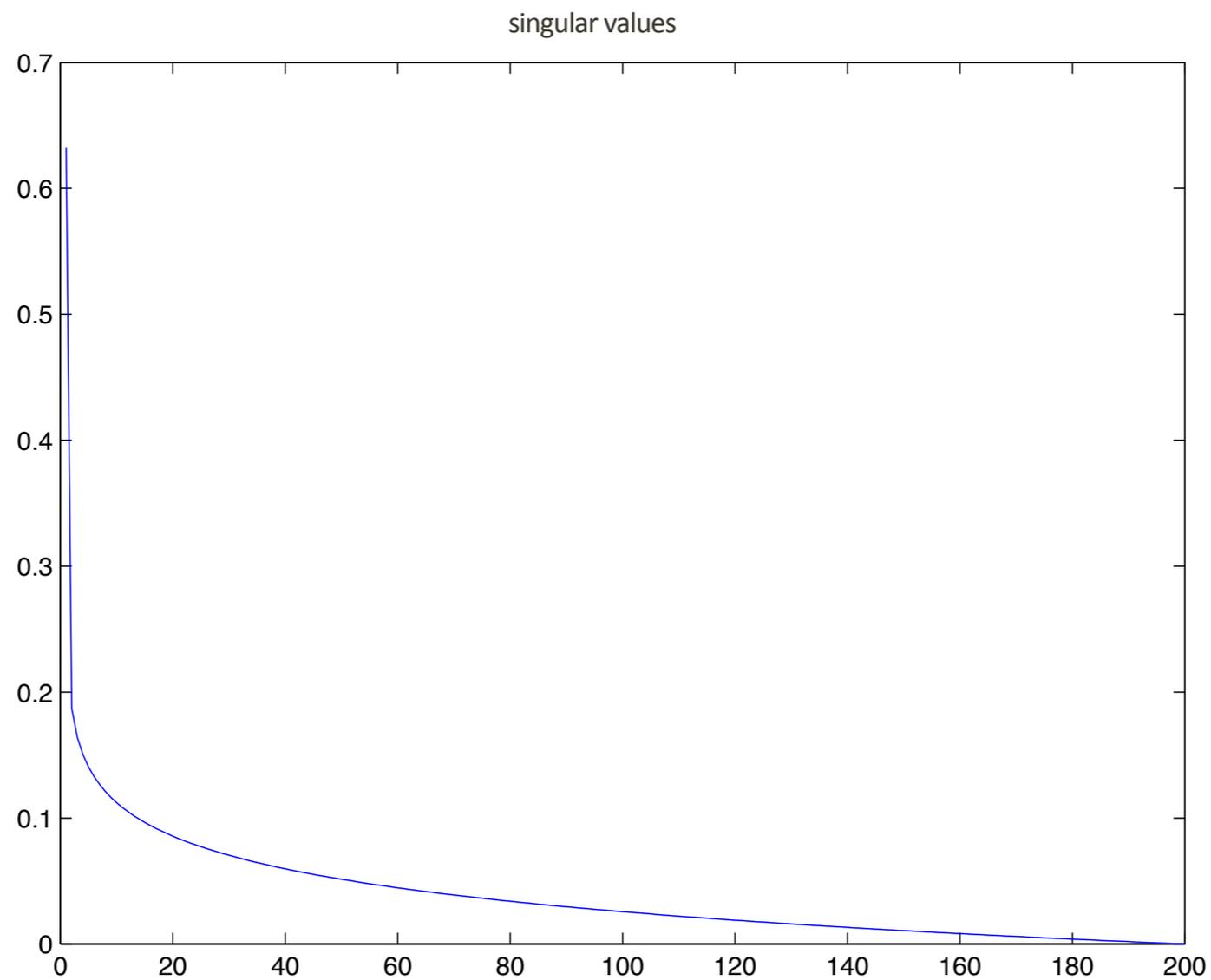
# SPGI1 vs PQNI1

(ill conditioned problem)

- What if our sensing matrix is not that nice, what if they have huge condition numbers?

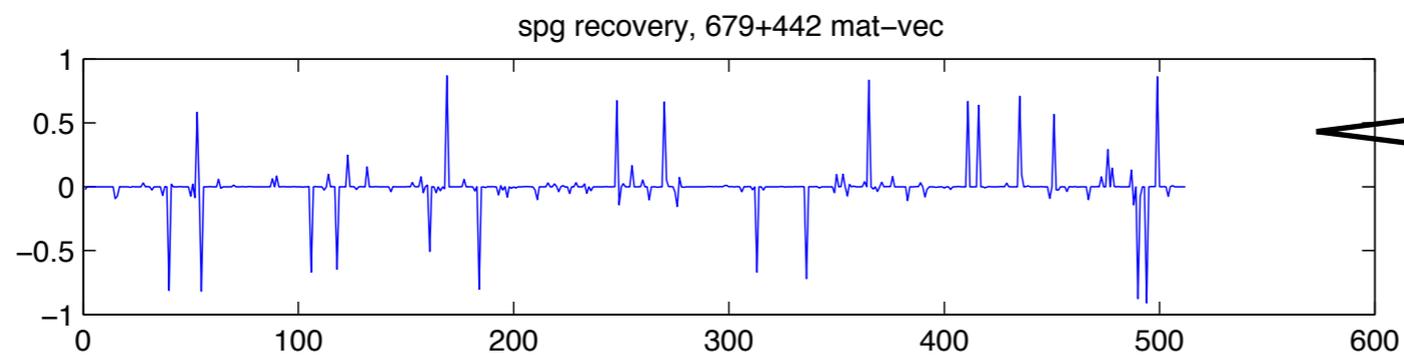
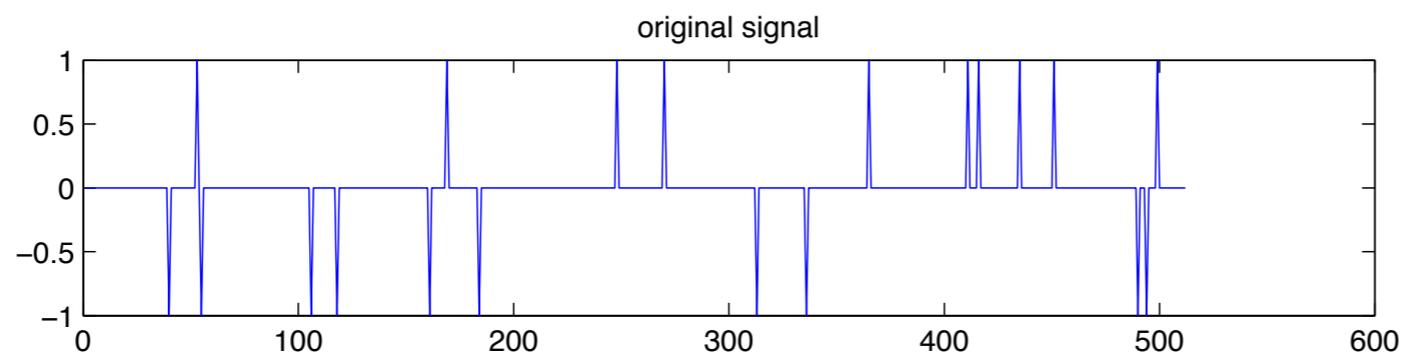
# SPGI1 vs PQNI1

(ill conditioned problem)

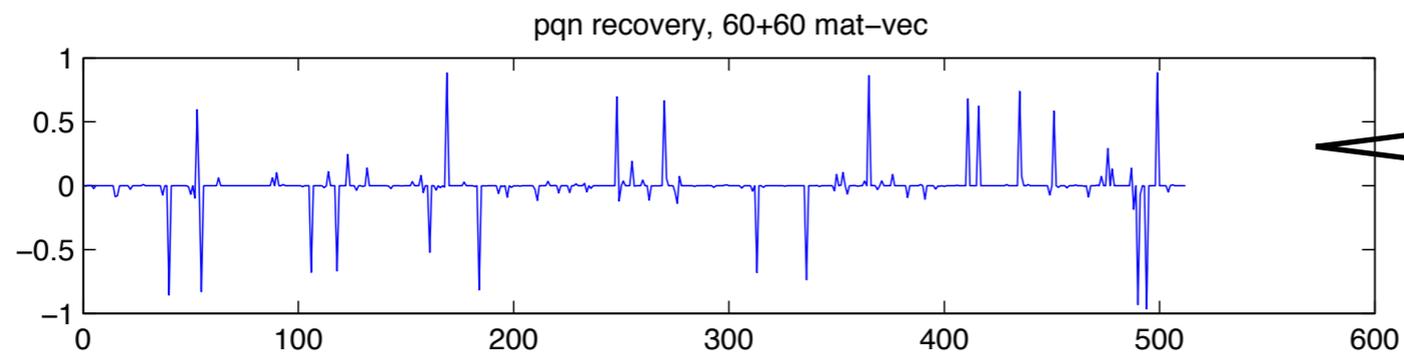


# SPG vs PQN

(ill conditioned problem)

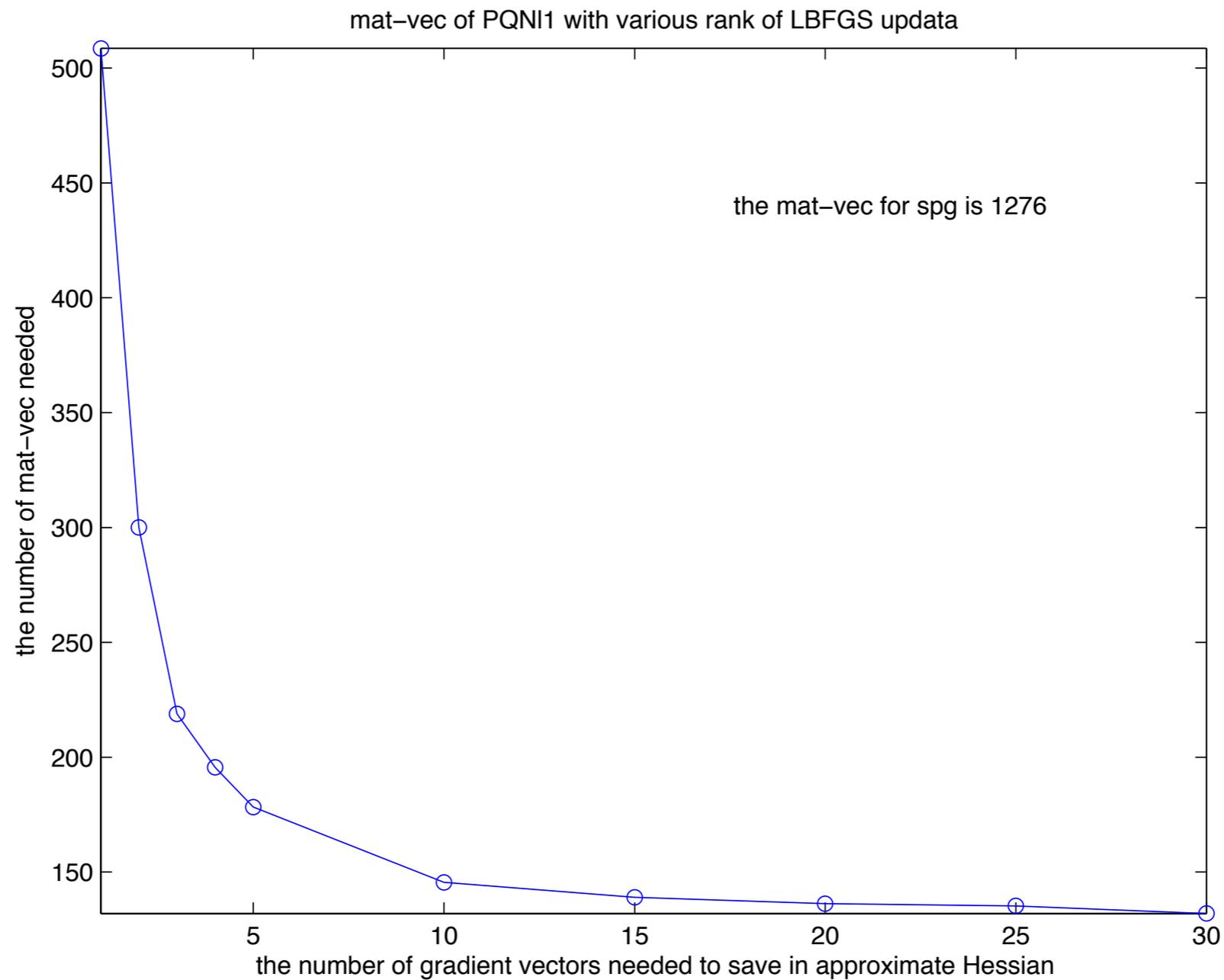


SPGL1 :  
692 + 442



PQNL1 :  
60 + 60

# Memory vs CPU time



# Observations

(from stylized numerical example)

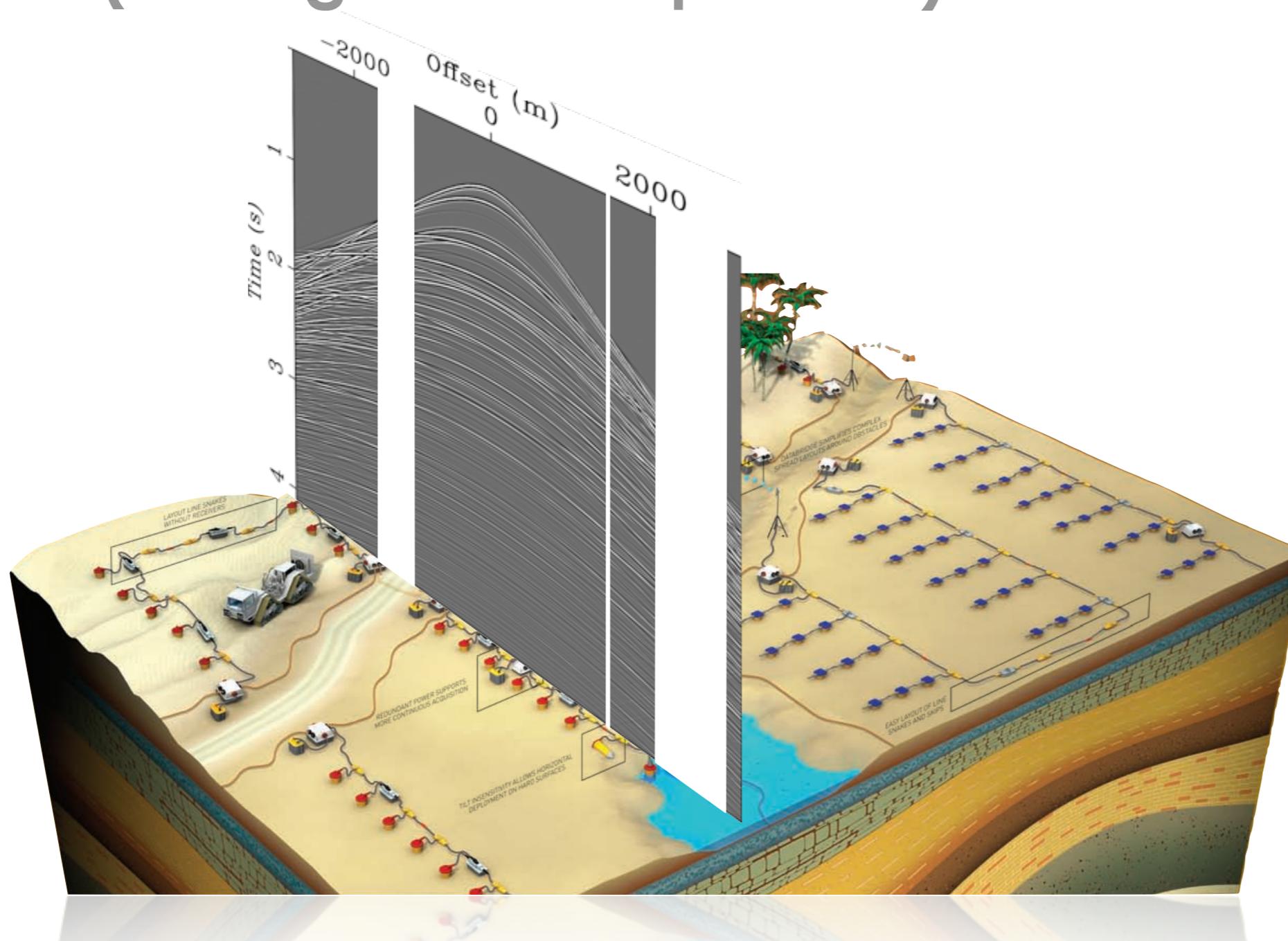
- PQN1 increases the convergence rate of SPG1
- For ill conditioned problem : PQN is more favorable
- Increase the approximate Hessian rank, the acceleration grows

# Experiments

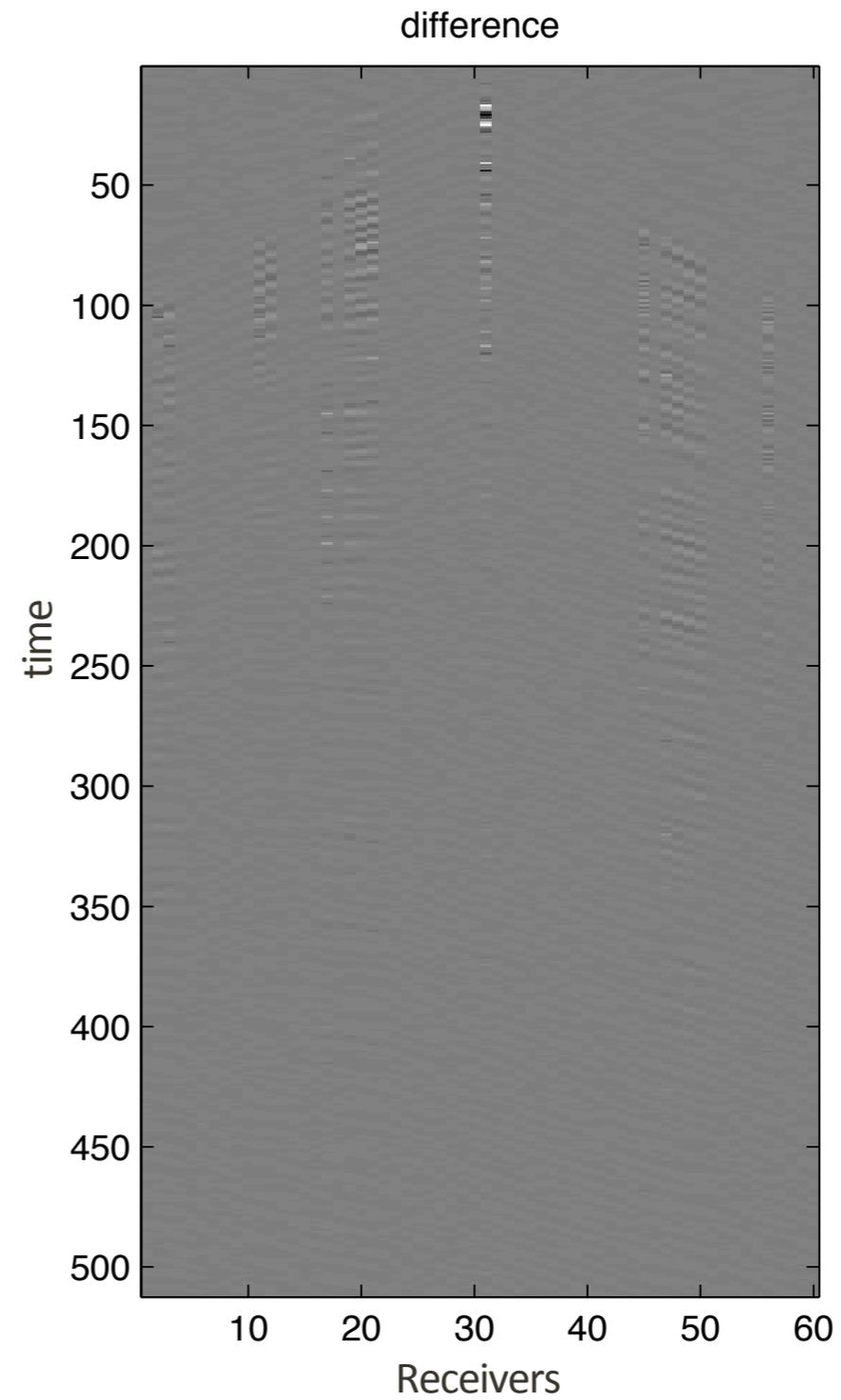
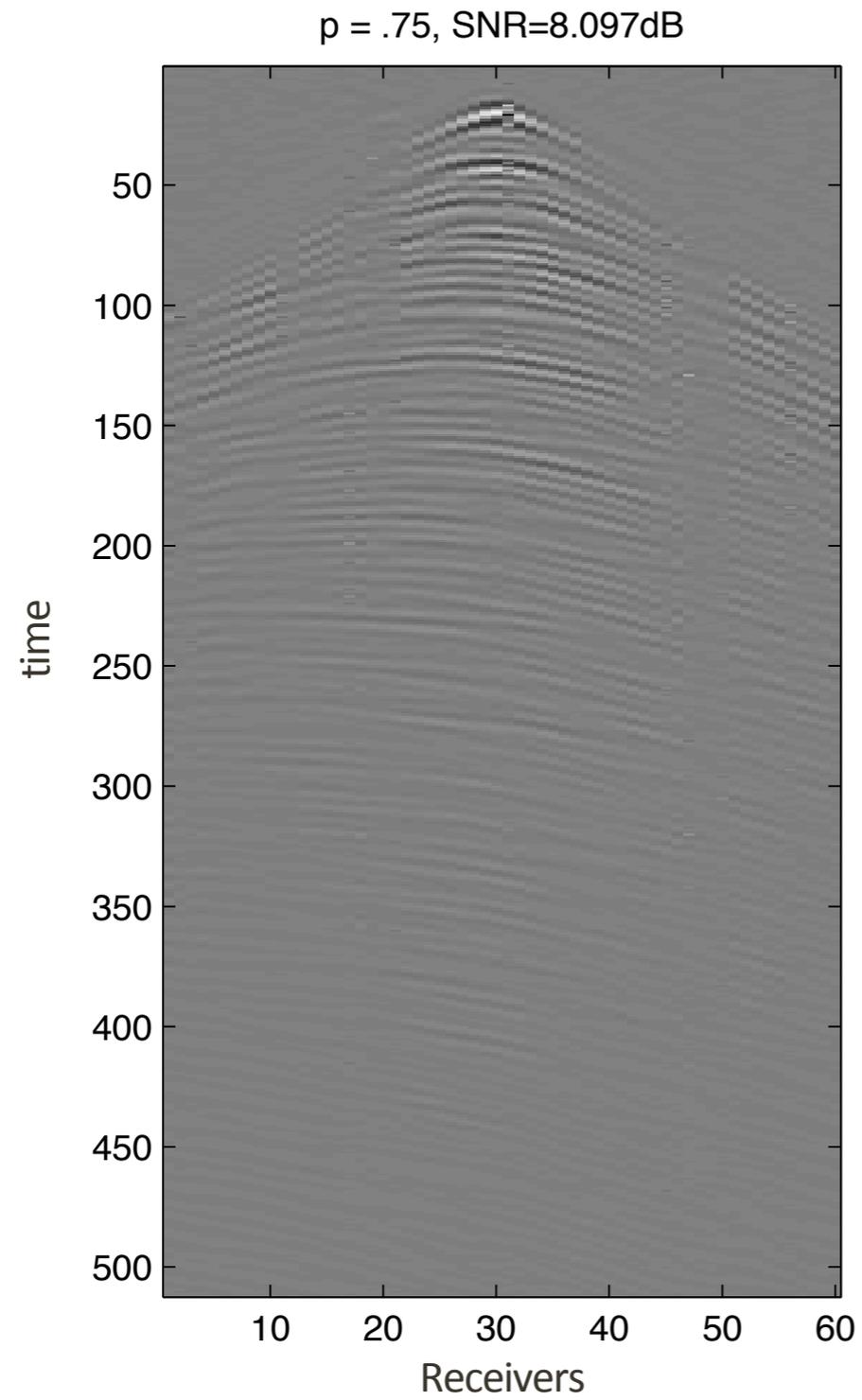
- General stylized numerical example
- Usage and benefits in seismic data acquisition, processing, and inversion

# PQNI1 in Seismic Acquisition

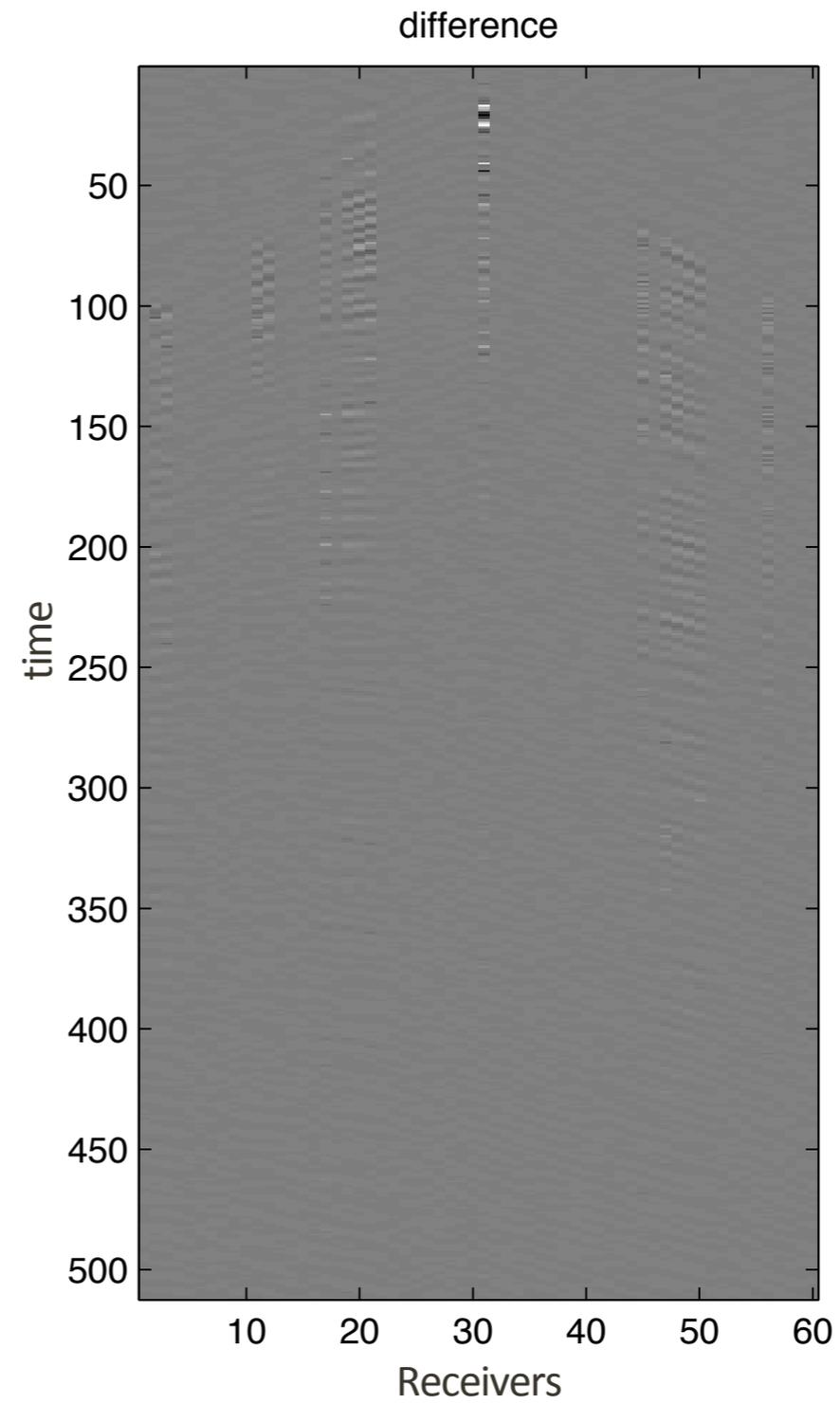
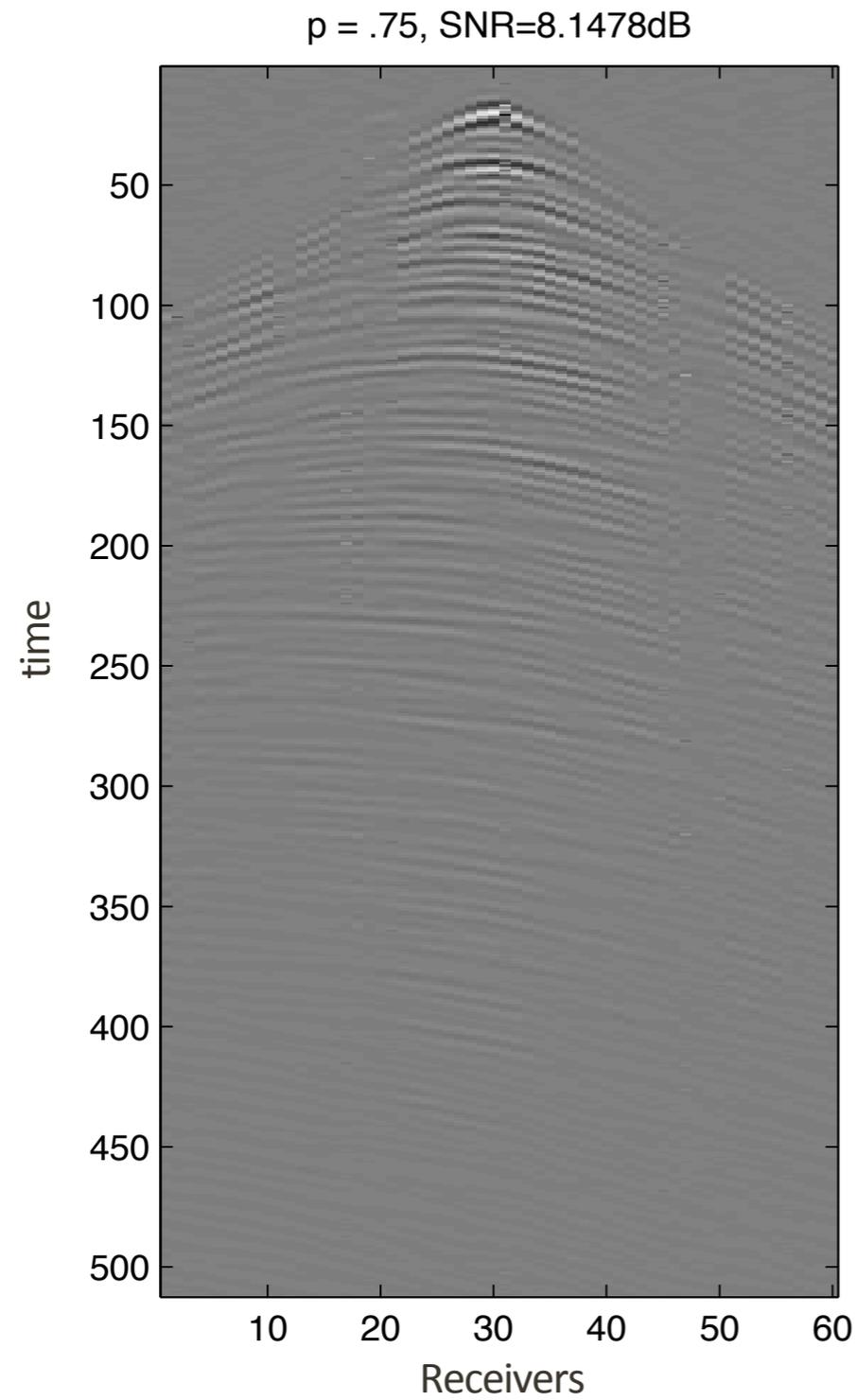
## (Missing Shots Interpolation)



# SPGI1 recovery



# PQNI1 recovery



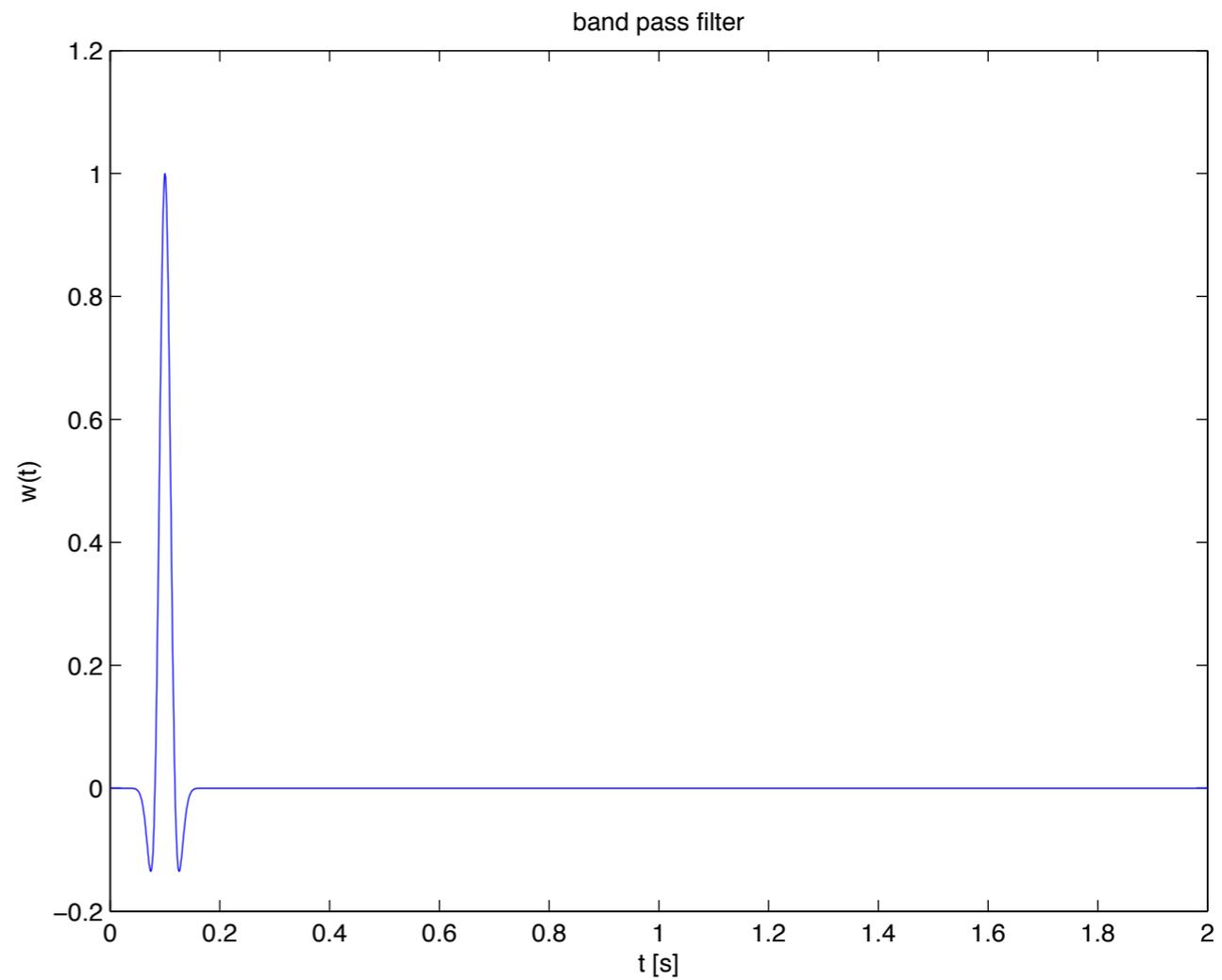
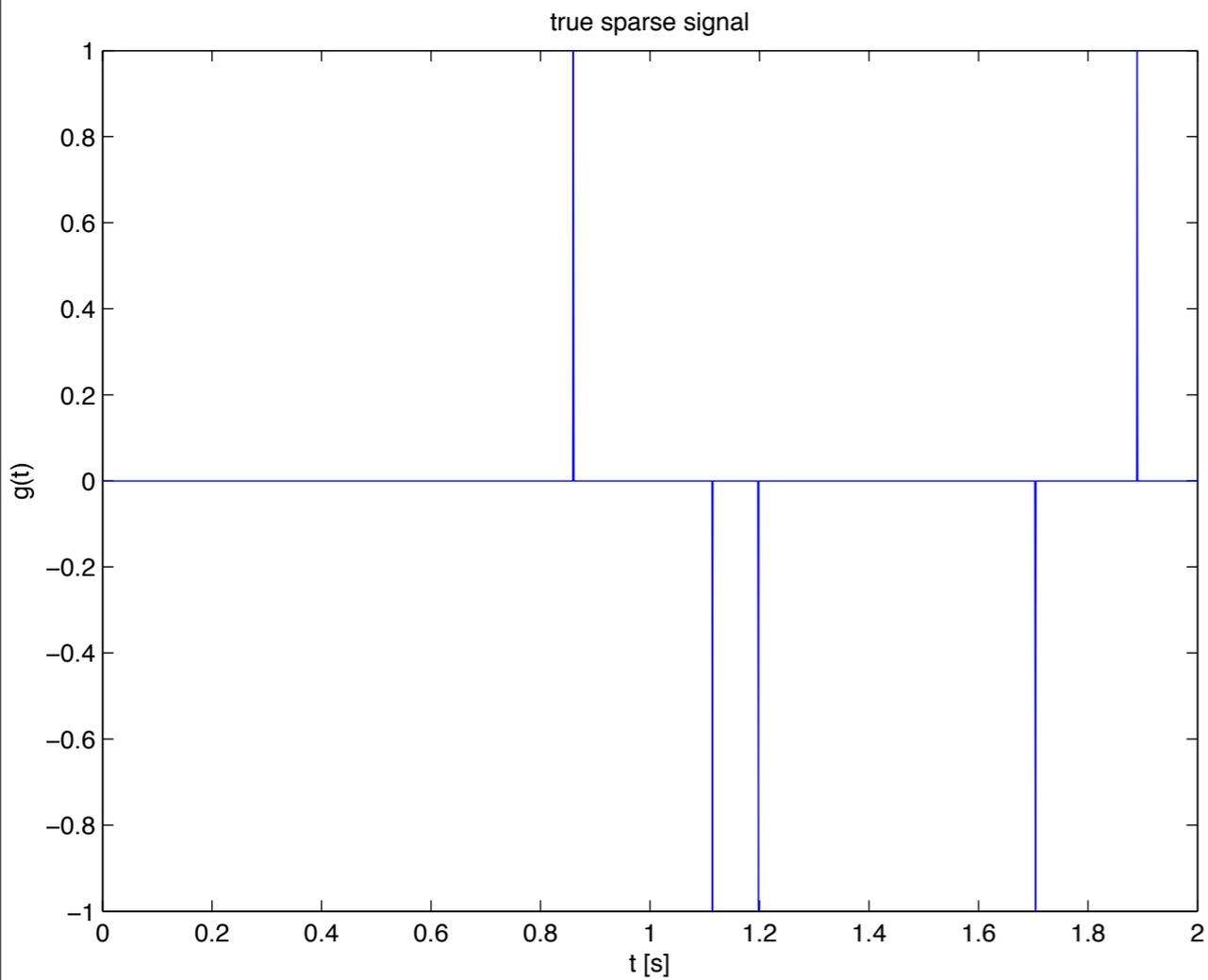
# SPG vs PQN

	ITERATION	MAT-VEC
SPGL1	300	423+247
PQNL1(5)	239	247+247
PQNL1(10)	214	222+222
PQNL1(20)	218	225+225
PQNL1(30)	217	225+225

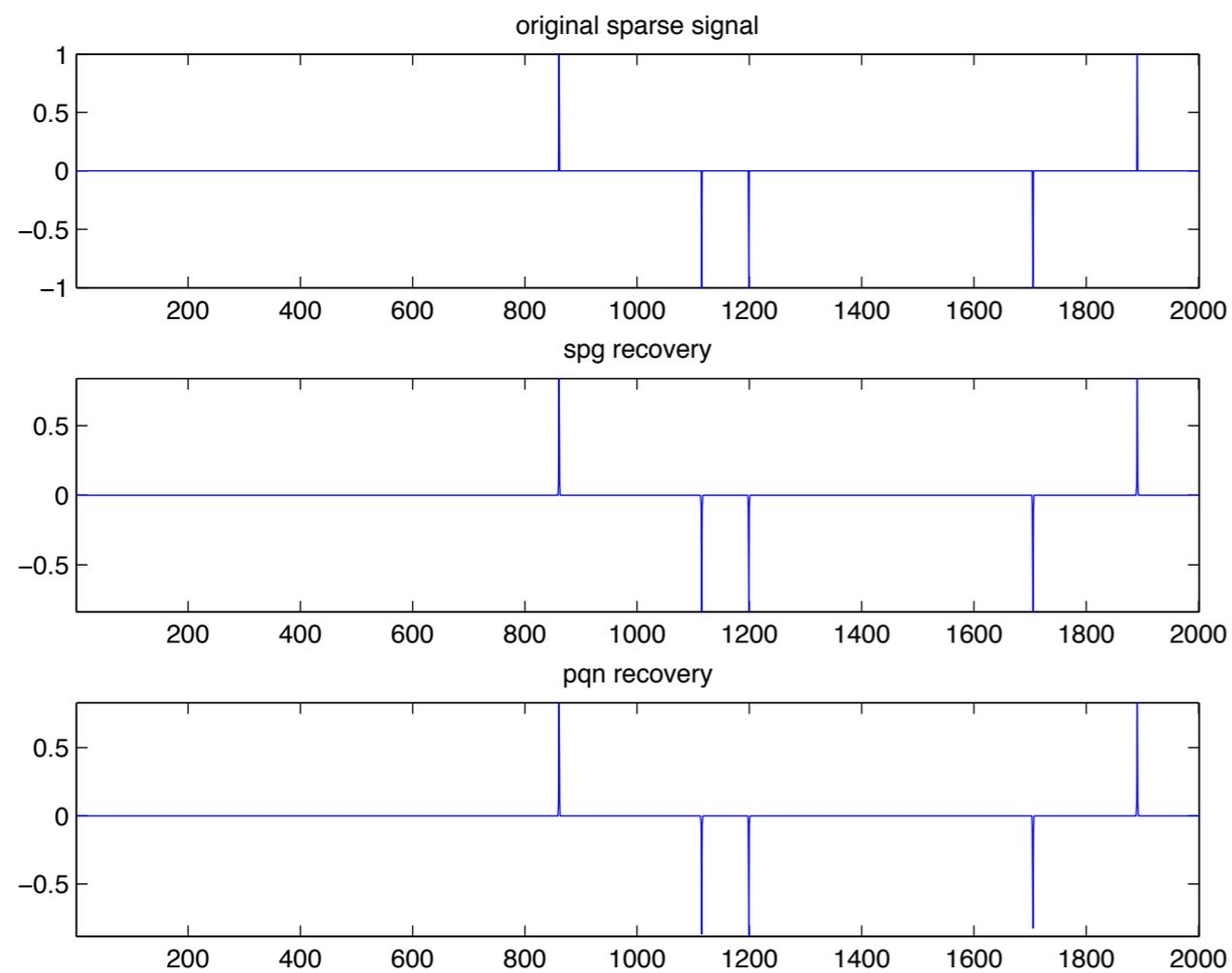
} Oscillation

# PQN1 in Data Processing

## (Deconvolution)



# SPGL1 vs PQNL1

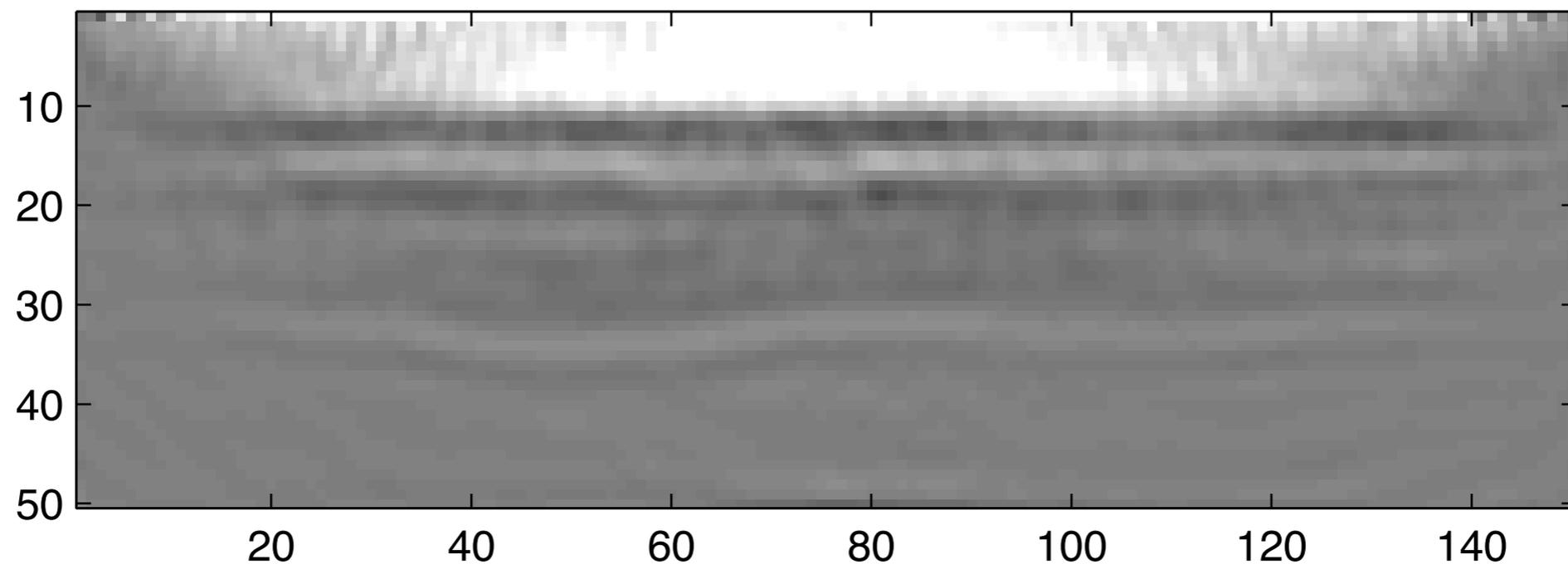


	ITERATION	MAT-VEC
<b>SPGL1</b>	1950	2907+1951
<b>PQNL1(5)</b>	820	822+822
<b>PQNL1(10)</b>	380	382+382
<b>PQNL1(20)</b>	370	372+372
<b>PQNL1(30)</b>	341	343+343

# PQNI1 in Seismic Imaging

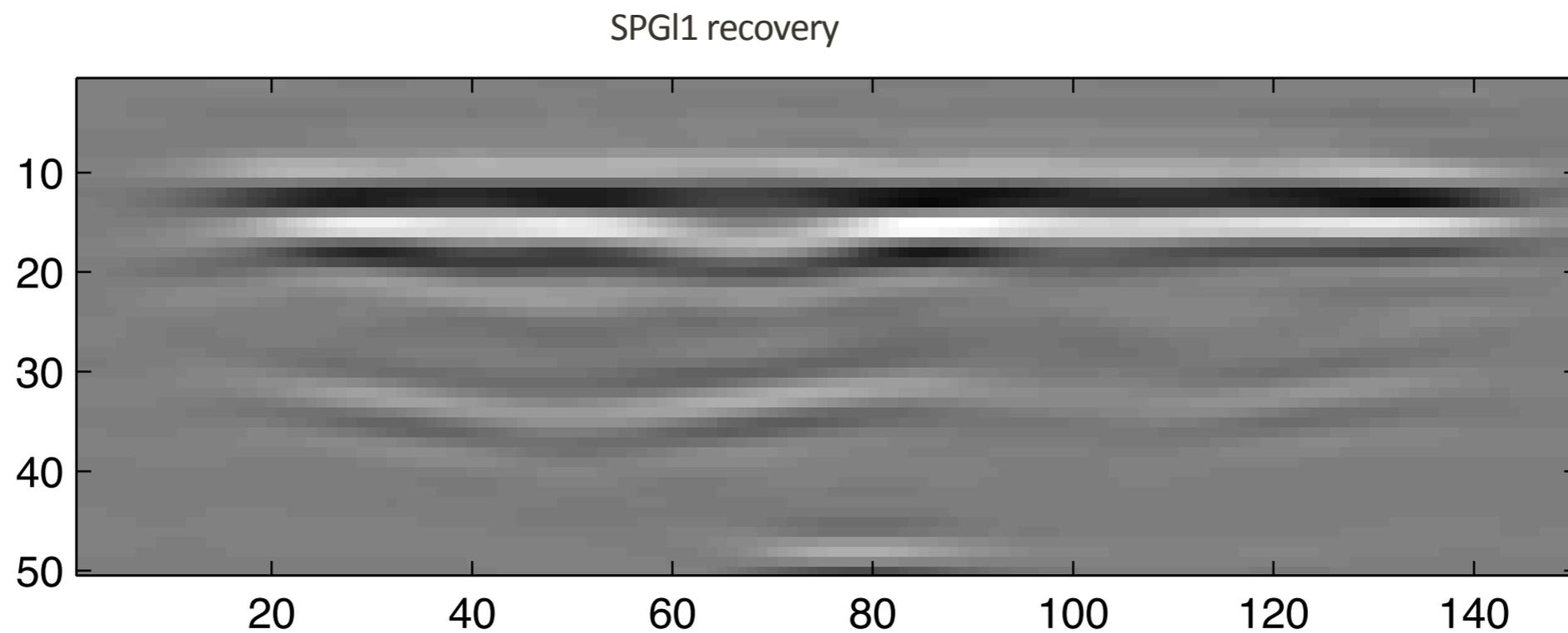
## (Reverse time migration)

Reverse time migration result



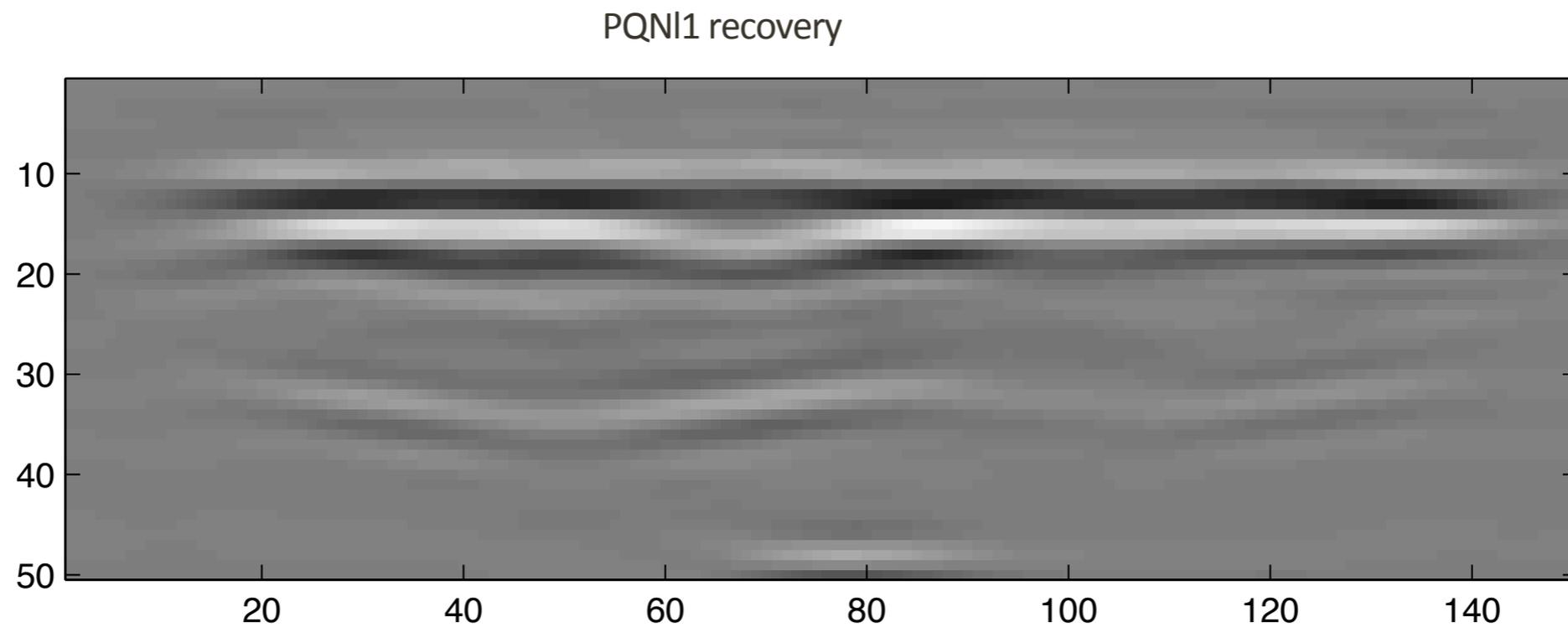
# PQNI1 in Seismic Imaging

(Reverse time migration)



# PQNI1 in Seismic Imaging

(Reverse time migration)



# PQNL1 in seismic imaging

(Reverse time migration)

	ITERATION	MAT-VEC	TIME(s)
SPGL1	100	156+101	39375
PQNL1(5)	36	45+45	9278.7
PQNL1(10)	35	44+44	7832.4
PQNL1(20)	35	44+44	6515.3
PQNL1(30)	35	44+44	6335.9

# Observation

(from seismic applications)

- PQN is a useful tool in various seismic applications
- Seismic applications with expensive `mat_vec` benefit more from PQN1
- Ill conditioned image space type problem is most favorable setting for PQN1

# Conclusion

- If the objective function is go from image space to data space
- or if memory is not an issue
- or if the problem is ill conditioned
  
- Consider PQNI1

## Future work

- Test and improve the robustness of PQNI1
- Further increase the convergence rate for PQNI1
- Promote the usage of PQNI1 into FWI

# Acknowledgements

- Mark Schmidt, Sasha, Tristan, Xiangli, Haneet
- SPGI1, SPOT, Curvelab
- BG compass dataset
- SLIM group

**SINBAD**



This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco.

# References

E. van den Berg and M. P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, 2008

M. Schmidt, E. van den Berg, M. Friedlander, and K. Murphy. Optimizing costly functions with simple constraints: A limited-memory projected quasi-newton algorithm. Conference on Artificial Intelligence and Statistics, 2009b.

M. Schmidt, G. Fung, and R. Rosales. Fast optimization methods for L1 regularization: A comparative study and two new approaches. European Conference on Machine Learning, 2007a.

Mark Schmidt, Graphical Model Structure Learning with l1-Regularization, 2010, PhD thesis.

Birgin, Martinez, Raydan, 2000, Nonmonotone spectral projected gradient methods on convex sets.

Byrd, Richard H.; Lu, Peihuang; Nocedal, Jorge; Zhu, Ciyou (1995). "A Limited Memory Algorithm for Bound Constrained Optimization".

Alen Andrew, Jianfeng Gao, Scalable training of L1 regularized loglinear models, ICML 2007

[http://www.google.ca/search?client=safari&rls=en&q=newtons+method+for+root+find&ie=UTF-8&oe=UTF-8&redir\\_esc=&ei=7MyQT5rpGeiViAKJzY2gAw](http://www.google.ca/search?client=safari&rls=en&q=newtons+method+for+root+find&ie=UTF-8&oe=UTF-8&redir_esc=&ei=7MyQT5rpGeiViAKJzY2gAw)

[http://en.wikipedia.org/wiki/Limited-memory\\_BFGS](http://en.wikipedia.org/wiki/Limited-memory_BFGS)

[http://en.wikipedia.org/wiki/Newton's\\_method\\_in\\_optimization](http://en.wikipedia.org/wiki/Newton's_method_in_optimization)